

J Vis (2011) 14:393–398
DOI 10.1007/s12650-011-0101-2

REVIEW PAPER

Seiichiro Morizawa · Koji Shimoyama ·
Shigeru Obayashi · Kenichi Funamoto ·
Toshiyuki Hayase

Implementation of visual data mining for unsteady blood flow field in an aortic aneurysm

Received: 16 June 2010 / Accepted: 10 July 2011 / Published online: 27 August 2011
© The Author(s) 2011. This article is published with open access at Springerlink.com

Abstract This study was performed to determine the relations between the features of wall shear stress and aneurysm rupture. For this purpose, visual data mining was performed in unsteady blood flow simulation data for an aortic aneurysm. The time-series data of wall shear stress given at each grid point were converted to spatial and temporal indices, and the grid points were sorted using a self-organizing map based on the similarity of these indices. Next, the results of cluster analysis were mapped onto the real space of the aortic aneurysm to specify the regions that may lead to aneurysm rupture. With reference to previous reports regarding aneurysm rupture, the visual data mining suggested specific hemodynamic features that cause aneurysm rupture.

Keywords Visual data mining · Self-organizing map · Aortic aneurysm · Wall shear stress

1 Introduction

Flow visualization using isosurfaces, volume rendering, etc. should help us to understand the flow fields from computational fluid dynamics (CFD) simulation data. However, even through the flow visualization, it is still difficult to understand the phenomena latent in the unsteady flow fields. This is because unsteady CFD simulations provide a huge amount of data, which consist of various fluid properties (density, velocity, pressure, etc.), given at each grid point at every time step. Therefore, most of the data users usually narrow down the data to limit size, which is acceptable for manual data processing based on their experience and/or intuition. Such a data processing approach may lead to the overlooking of universal information about unsteady flow fields.

To systematically treat unsteady CFD data, feature extraction techniques are used together with flow visualization (e.g., Jiang et al. 2002). They introduce the technique that can automatically extract vortex features by searching for vortex core regions. In general, however, there are other characteristic flow features, such as shock waves, shear layers, and separation lines. In addition, flow unsteadiness should be extracted as an additional feature for unsteady flow fields. Consequently, to reveal universal information about unsteady flow fields, various flow features and their relations need to be examined in a comprehensive manner.

Our research group has been proposing the use of “visual data mining,” in which an additional process called “data mining” is integrated with feature extraction and visualization, for understanding unsteady flow fields. Data mining is an automated process to find characteristic rules and/or patterns latent in high-dimensional data, which are too complicated to be interpreted manually. Our ideas are to analyze multiple

S. Morizawa (✉) · K. Shimoyama · S. Obayashi · K. Funamoto · T. Hayase
Institute of Fluid Science, Tohoku University, 2-1-1 Katahira, Aoba-ku, Sendai 980-8577, Japan
E-mail: morizawa@edge.ifs.tohoku.ac.jp
Tel.: +81-22-2174438
Fax: +81-22-2174438

features extracted from unsteady flow fields through the data mining process, and then to visualize characteristic patterns of unsteady flow phenomena based on the data mining results. Thus, the visual data mining is expected to support a comprehensive understanding of unsteady flow fields.

To date, our research group has demonstrated the implementation of visual data mining for unsteady CFD data of supersonic airflow to determine the relations between shock wave oscillation and vortex shedding (Shibasaki and Obayashi 2006). They successfully extracted and visualized specific information about unsteady supersonic flow, which agreed with empirical fluid dynamic knowledge, but it did not achieve the acquisition of brand-new information about significant flow phenomena and mechanisms. Based on such background, this study aims to achieve an earnest acquisition of unknown flow information through visual data mining. This study deals with the unsteady blood flow simulation data for an aortic aneurysm as one of the application example; it performs the visual data mining to examine the relations between blood flow features and aneurysm rupture, and to provide significant information about the blood flow phenomena and mechanisms, which is useful for predicting aneurysm rupture.

2 Hemodynamics related to aneurysm rupture

Several previous studies have revealed that wall shear stress (WSS) plays an important role in the occurrence of circulatory diseases, such as arteriosclerosis and aneurysm rupture (Caro et al. 1971). This is because temporal variation in WSS affects the hemodynamic characteristics of endothelial cells. However, aneurysm rupture is considered to show different physical features depending on the situation.

There have been conflicting reports suggesting that low or high WSS causes aneurysm rupture. The concept of low WSS suggests that aneurysm rupture tends to occur in regions where the peak value of WSS during the cardiac cycle is reduced (Shojima et al. 2004), because such WSS conditions are apt to suppress blood circulation, cause blood vessel wall degeneration, and necrotize vascular cell. In contrast, the concept of high WSS, which suggests that aneurysm rupture may occur at regions with a large WSS peak value (Hassan et al. 2004), also seems plausible as high WSS is directly linked to a mechanical contribution to aneurysm rupture. This paper reported an example corresponding to the high WSS concept in which blood flow is separated from the parent artery and hits on the inside wall of the aneurysm, and therefore, WSS becomes high locally at this position and may lead to aneurysm rupture.

3 Visual data mining procedure

Figure 1 shows a flowchart of the present visual data mining procedure. This procedure consists of four steps: the first step inputs the time-series data obtained from unsteady flow simulation. The second step extracts temporal indices from the time-series data for each grid point. Based on the similarity of the temporal and spatial indices, the third step divides a set of grid points into several clusters by self-organizing map (SOM) (Kohonen 1995). The final step maps the clustering results onto a real space for visualizing the temporal and spatial characteristics latent in the time-series data. These procedures enable us to examine complex characteristics of unsteady flow fields easily. Details of each step are described in Sects. 3.1–3.4

3.1 Input data

The input data to be analyzed in this study were obtained from the unsteady blood flow simulation for an aortic aneurysm reported previously (Funamoto et al. 2008). Figure 2 shows the aortic aneurysm considered in this simulation. Figure 2a shows the whole shape layout of the aortic aneurysm, which was reconstructed from a patient's computed tomography image using the software Mimics and Magics. The blood flow goes out of the heart through ascending aorta, aortic arch (upper curved region), descending aorta, and abdominal aorta, and afterward it reaches the lower body. Figure 2b shows the simulation domain in the descending aorta section with an aneurysm.

These data contain the time-series data of WSS vectors in a three-dimensional Cartesian coordinate system. The WSS data are given at each of 3,631 grid points on the aorta surface during a cardiac cycle of 0.98 s (98 time steps with a step size of 0.01 s). Unlike typical flow simulations, upstream and downstream boundary conditions were not set to be constant in the blood flow simulation. Prior to this simulation, unsteady blood flow in the whole descending aorta was simulated by FLUENT. These results were then

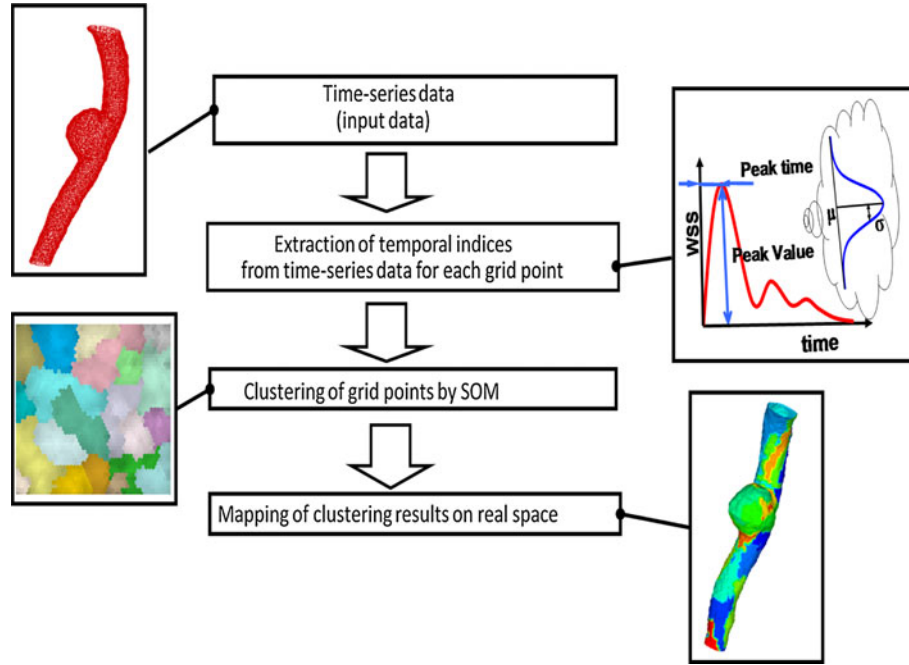


Fig. 1 Flowchart of visual data mining procedure

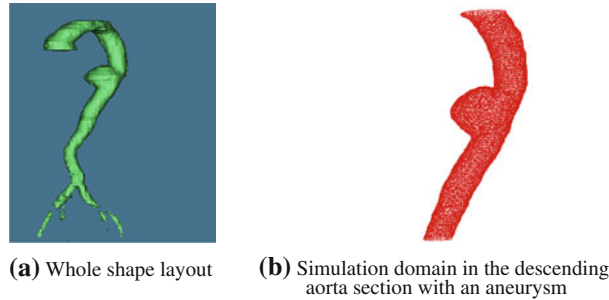


Fig. 2 Aortic aneurysm

applied to the unsteady boundary conditions for simulation in the partial descending aorta near the aneurysm. Therefore, these data did not lack accuracy even near the upstream and downstream boundaries. Notes that these data have already been validated experimentally in terms of velocity fields by comparison with actual measurement using Doppler velocimetry in the reference (Funamoto et al. 2008).

3.2 Extraction of temporal indices

For each grid point, five temporal index values ($wssmx$, $tmmx$, $wssav$, CV , and OSI) are extracted from the WSS time-series data. “ $wssmx$ ” is the maximum peak value of WSS in a cardiac cycle, and “ $tmmx$ ” is the time when $wssmx$ occurs. These two index values represent the attenuation and delay of cardiac pulse. “ $wssav$ ” is the time-averaged value of WSS during a cardiac cycle. “ CV ” (coefficient of variation) is the ratio of the standard deviation (σ) to the average (μ) of WSS. CV represents the “magnitude” unsteadiness of WSS value at each grid point. On the other hand, “ OSI ” (oscillatory shear index) can be used to quantify the “directional” unsteadiness of the WSS vector at each grid point as follows (He and Ku 1996):

$$OSI = \frac{1}{2} \left\{ 1 - \frac{\left| \int_0^T wss_i dt \right|}{\int_0^T |wss_i| dt} \right\} \quad (1)$$

where wss_i is an instantaneous WSS vector at time t , and T is the time period of the cardiac cycle.

3.3 Cluster analysis by self-organizing map (SOM)

Cluster analysis aims to group the dataset into several subsets (called clusters), so that data points in the same cluster are similar in terms of data dimension values. In this study, we performed cluster analysis for a set of 3,631 grid points by SOM, based on the similarity of the total of eight index values: five temporal indices (*wssmx*, *tmnx*, *wssav*, *CV*, and *OSI*) and three spatial indices (*x*, *y*, and *z* coordinates) assigned to each grid point.

SOM is a feedforward-type neural network model with an unsupervised learning algorithm, as illustrated in Fig. 3. Neurons in the input layer of SOM are associated with the input data vectors as:

$$f^i = [f_1^i, f_2^i, \dots, f_m^i]^T \quad (i = 1, 2, \dots, N) \quad (2)$$

On the other hand, neurons in the output layer are arranged with two-dimensional rectangular or hexagonal mesh topology, and associated with the weight vectors as:

$$w^j = [w_1^j, w_2^j, \dots, w_m^j]^T \quad (j = 1, 2, \dots, L) \quad (3)$$

where m is the number of input vector dimensions, N is the number of neurons in the input layer (equivalent to the number of input data points), and L is the number of neurons in the output layer. Note that the weight vectors have the same m dimensions as the input vectors and are randomly assigned before the learning process. The learning algorithm of SOM starts with finding the best-matching unit w^{c_i} , which is the closest weight vector to each input vector f^i as follows:

$$\|f^i - w^{c_i}\| = \min_j \|f^i - w^j\| \quad (j = 1, 2, \dots, L) \quad (4)$$

Once the best-matching units are determined for all input vectors, the weight adjustments are performed not only for the best-matching units but also for the neighbors. The adjustments depend on the distance (similarity) between input vectors and weight vectors. The weight vector w^j is adjusted to w_{adj}^j as follows:

$$w_{\text{adj}}^j = \sum_{i=1}^N h_{j c_i} f^i / \sum_{i=1}^N h_{j c_i} \quad (j = 1, 2, \dots, L) \quad (5)$$

where $h_{j c_i}$ is defined by the following Gaussian-like function:

$$h_{j c_i} = \exp\left(-\frac{\mathbf{d}_{j c_i}^2}{\mathbf{r}_t^2}\right) \quad (6)$$

where $\mathbf{d}_{j c_i}$ denotes the Euclidean distance between the neuron w^j and the best-matching unit w^{c_i} on the two-dimensional map in the output layer, and r_t denotes the neighborhood radius, which decreases with iteration of the learning processes. Repeating the learning process, weight vector distribution becomes smooth not only locally, but also globally on the two-dimensional map.

The SOM performs nonlinear projection of high-dimensional input dataset onto a two-dimensional map, so that a sequence of close data points in the original high-dimensional space will result in a sequence of neighboring neurons in the two-dimensional map. Thus, the SOM makes it easy to group the high-dimensional data into clusters on the two-dimensional map. In addition, the SOM enables the high-dimensional data to be visualized in a two-dimensional form while preserving their own features. Therefore, comparing the SOM images colored by the data dimension values, users can interpret the correlation patterns among the dimensions qualitatively.

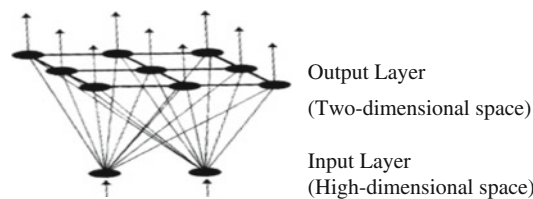


Fig. 3 Basic structure of SOM

3.4 Mapping of clustering results onto real space

The clusters divided by SOM show different unsteady flow field characteristics from each other. To visualize these characteristics in an intuitive manner, clustering results, such as cluster IDs, are mapped onto the real space of the aorta surface. The mapping results help us to visually distinguish real-space regions with different temporal flow features.

4 Results

Figure 4 shows the visual data mining results, in which the set of 3,631 grid points is divided into 21 clusters by the SOM based on the similarity of spatial index values (x , y , and z) and temporal index values ($wssmx$, $tmmx$, $wssav$, CV , and OSI). Figure 4a shows the SOM images colored by each index value, Fig. 4b shows the SOM image divided into clusters, and Fig. 4c shows the contour plots of SOM cluster IDs, which are mapped onto the real space of the aortic aneurysm surface. Here, note that the data given at a grid point maintains the same position in all SOM images. In Fig. 4b, c, numbers 1–5 are assigned to the clusters, each of which involves some features concerned with aneurysm rupture and are discussed below.

Comparison of the color patterns among the SOM images in Fig. 4a indicates that $tmmx$ and $wssmx$ show a similar tendency, such that the index value changes markedly between the area corresponding to the aneurysm (denoted as clusters 1, 2, and 3 in Fig. 4b, c) and the others. In addition, cluster 5 corresponding to the edge of aneurysm is located far from cluster 2 in the SOM image as shown in Fig. 4b, although these clusters are adjacent in the real space as shown in Fig. 4c. This means that WSS temporal characteristics change markedly between the aneurysm and its surroundings.

Hereafter, the relations between WSS temporal characteristics in an aortic aneurysm and aneurysm rupture are examined based on the concepts reported in previous studies as described in Sect. 2. First, the low WSS concept for aneurysm rupture applies to cluster 2 corresponding to the location inside the aneurysm, because this cluster has a low value of $wssmx$. With regard to other index values, cluster 2 includes low CV and low OSI , which indicate the blood flow conditions with small temporal variations in terms of WSS magnitude and direction. From the standpoint of fluid dynamics, these results seem consistent with the actual condition that blood circulation becomes inactive. In addition, cluster 2 has a much larger value of $tmmx$ than any other cluster. This result means that the WSS time-series data at the locations far

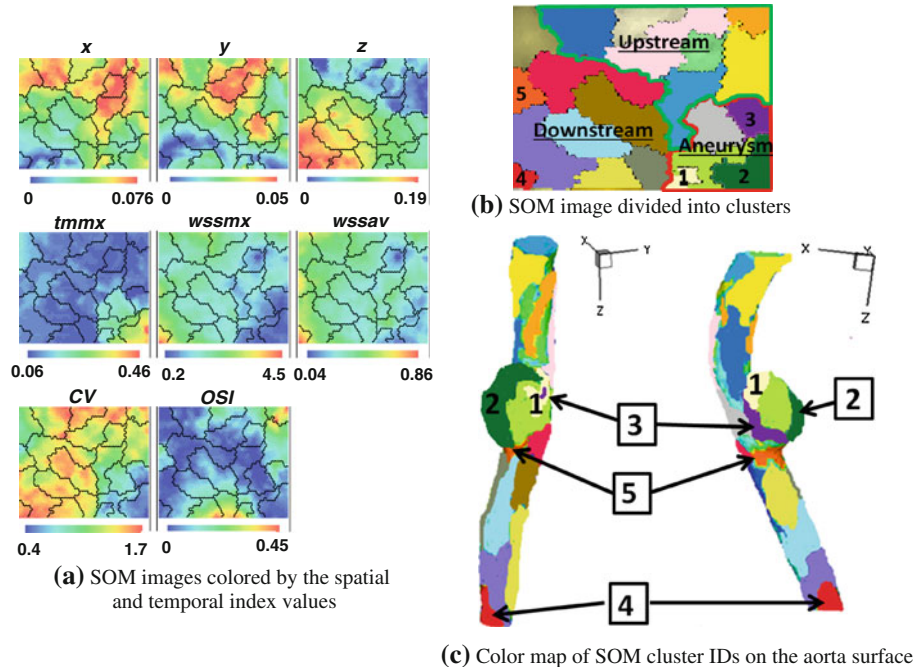


Fig. 4 Visual data mining results

from the aneurysm have strong peaks that synchronize with the blood pulsation by the heart, while the data inside the aneurysm have gentle and delayed peaks due to inactive blood circulation, i.e., the blood flow inside the aneurysm is insensitive to the blood pulsation. Therefore, the above discussions based on the low WSS concept suggest that aneurysm rupture tends to occur wherever blood flow is insensitive to heart pulsation.

On the other hand, the high WSS concept for aneurysm rupture applies to cluster 4 because this cluster has high *wssmx*. With regard to other index values, cluster 4 has high *CV* and low *OSI*. These index values indicate the flow conditions with high WSS and small temporal variation of WSS direction, i.e., high WSS always acts on the aorta surface in a same direction. Under such conditions, the blood vessel wall tends to yield by accumulating strain energy. However, it should be noted that the input data used in this study were obtained from the unsteady blood flow simulation, which assumed the blood vessel to be rigid without strain energy. Cluster 4 corresponds to the downstream end of the aorta. Therefore, the present results indicated that the high WSS concept is inappropriate to identify the regions where aneurysm rupture will occur, while the low WSS concept is appropriate for this purpose.

5 Conclusions

Visual data mining of the unsteady blood flow simulation data for an aortic aneurysm was performed to examine the relations between the features of WSS and aneurysm rupture. The visual data mining extracted spatial and temporal indices from time-series WSS data given at each grid point, and then clustered the grid points using a SOM based on the similarity of the index values. The results identified the critical regions where aneurysm rupture may occur with reference to the existing reports. Then, the visual data mining revealed the unsteady hemodynamic features, which seemed to be closely related to aneurysm rupture. The etiological factors of aneurysm rupture are attributable to the regions in which circulation of blood is inactive. Consequently, these results confirmed the superior capability of visual data mining as one of the systematic methodologies to provide important information about hemodynamic features as well as for prediction of aneurysm rupture in an aortic aneurysm.

A key issue in this paper is that visual data mining has been established for unsteady CFD data, and suggested specific hemodynamic features, which seem possible to cause aneurysm rupture, from a single case of blood flow simulation data. However, it is undeniable that the number of blood flow simulation cases considered here is insufficient to verify the relations between WSS and aneurysm rupture from the standpoint of medical science. Therefore, further implementation of visual data mining in other cases and medical verification of those results should be considered as future works.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Caro CG, Fitzgera JM, Schroter RC (1971) Atheroma and arterial wall shear? Observation, correlation and proposal of a shear dependent mass transfer mechanism for atherogenesis. *Proc R Soc Lond B* 177:109–159
- Funamoto K, Hayase T et al (2008) Numerical experiment for ultrasonic-measurement-integrated simulation of three-dimensional unsteady blood flow. *Ann Biomed Eng* 36:1383–1397
- Hassan T, Timofeev EV et al (2004) Computational replicas: anatomic reconstructions of cerebral vessels as volume numerical grids at three-dimensional angiography. *AJNR* 25:1356–1365
- He X, Ku DN (1996) Pulsatile flow in the human left coronary artery bifurcation: average conditions. *J Biomech Eng* 118:74–82
- Jiang M, Machiraju R, and Thompson D (2002) A novel approach to vortex core region detection. In: *Proceedings of Joint Eurographics and IEEE/TCVG symposium on visualization*, Vienna, Austria, October 2002, pp 217–226
- Kohonen T (1995) *Self-organizing maps*. Springer, Berlin
- Shibasaki T, and Obayashi S (2006) Construction of a data mining method for unsteady CFD results. In: *Proceedings of the JSME 19th computational mechanics conference*, Nagoya, Japan, 3–5 November, pp 431–432 (in Japanese)
- Shojima M, Oshima M et al (2004) Magnitude and role of wall shear stress on cerebral aneurysm: computational fluid dynamic study of 20 middle cerebral artery aneurysms. *Stroke* 35:2500–2505